P8131 HW8

1. Health

(a) Evaluate the bivariate, cross-sectional relationship between randomized group assignment and participants health self-rating at the time of randomization.

```
health <- readxl::read_excel("HW8-HEALTH.xlsx", col_names = TRUE) %>%
  janitor::clean_names() %>%
  mutate(id = as.factor(id),
         txt = as.factor(txt),
         health = factor(health, levels = c("Poor", "Good")),
         time = factor(time, ordered = TRUE),
         agegroup = as.factor(agegroup))
head(health) # data is sorted
```

```
## # A tibble: 6 x 5
##   id    time  txt          health agegroup
##   <fct> <ord> <fct>        <fct>  <fct>
## 1 101   1     Intervention Good   15-24
## 2 101   2     Intervention Good   15-24
## 3 101   3     Intervention Good   15-24
## 4 101   4     Intervention Good   15-24
## 5 102   1     Control      Poor   15-24
## 6 102   2     Control      Poor   15-24
```

Cross-sectional table:

		TXT			
		Control		Intervention	
		Count	Percentage %	Count	Percentage %
Good	15-24	10	52.63	9	47.37
	25-34	9	60.00	6	40.00
	35+	1	50.00	1	50.00
	Total	20	55.56	16	44.44
Poor	15-24	11	55.00	9	45.00
	25-34	7	36.84	12	63.16
	35+	3	60.00	2	40.00
	Total	21	47.73	23	52.27
Total	15-24	21	53.85	18	46.15
	25-34	16	47.06	18	52.94
	35+	4	57.14	3	42.86
	Total	41	51.25	39	48.75

Figure 1: Cross-Sectional table for randomized group assignment and participants health self-rating at the time of randomization

Interpret and discuss these findings:

Equal percentage

(b) Perform a longitudinal data analysis across all study follow-up visits

```
health.new <- subset(health, time > "1")
baseline <- rep(subset(health, time == "1"))
# make time 1 as another covariate: baseline
#113, 119, 123... only has 3 observations
add_baseline <- function(id){
  outcome = baseline$health[baseline$id == id]
  as.character(outcome)
}
health.new$baseline <- map(health.new$id, add_baseline) %>%
  factor(levels = c("Poor", "Good"))
health.new$nhealth <- as.numeric(health.new$health == "Good") # 1=good, 0=poor
health.new$postM <- (as.numeric(health.new$time)-1)*3 # month post randomization
# fit with unstructured correlation structure
health.new <- as.data.frame(health.new)
gee.health <- gee(nhealth ~ postM + txt + agegroup + baseline, data = health.new,
                  family = "binomial",
                  id = id,
                  corstr = "unstructured",
                  scale.fix = FALSE) # scale parameter is phi (over dispersion)
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
## running glm to get initial regression estimate
```

```
##     (Intercept)           postM txtIntervention   agegroup25-34     agegroup35+
##     -1.60936174      0.04404073      1.99778060      1.19586376      1.39542710
##    baselineGood
##      1.71129313
```

```
summary(gee.health)
```

```
##
##  GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
##  gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
##  Link:                      Logit
##  Variance to Mean Relation: Binomial
##  Correlation Structure:     Unstructured
##
## Call:
## gee(formula = nhealth ~ postM + txt + agegroup + baseline, id = id,
##     data = health.new, family = "binomial", corstr = "unstructured",
##     scale.fix = FALSE)
##
## Summary of Residuals:
##         Min          1Q      Median          3Q         Max
## -0.98120150 -0.18801168  0.09128879  0.17516123  0.83424138
##
##
```

```
## Coefficients:
##                    Estimate Naive S.E.    Naive z Robust S.E.  Robust z
## (Intercept)     -1.76899848 0.66103146 -2.6761184  0.57887503 -3.055925
## postM            0.05100277 0.06725101  0.7583941  0.07024227  0.726098
## txtIntervention  2.09950310 0.60087382  3.4940832  0.53792705  3.902951
## agegroup25-34    1.35098483 0.59300433  2.2782040  0.50386082  2.681266
## agegroup35+      1.41166003 0.98252381  1.4367693  0.78644380  1.794992
## baselineGood     1.81448641 0.60333504  3.0074276  0.51044100  3.554743
##
## Estimated Scale Parameter:  1.516997
## Number of Iterations:  5
##
## Working Correlation
##           [,1]      [,2]      [,3]
## [1,] 1.0000000 0.1743007 0.5809889
## [2,] 0.1743007 1.0000000 0.2049833
## [3,] 0.5809889 0.2049833 1.0000000
```

(c) Fit a generalized linear mixed effects model with subject-specific random intercepts

GLMM

```
glmm <- glmer(nhealth  ~ postM + txt + agegroup + baseline + (1 | id),
                  family = 'binomial', data = health.new)
summary(glmm) # pay attention to: random effects, fixed effects,
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: nhealth ~ postM + txt + agegroup + baseline + (1 | id)
##    Data: health.new
##
##      AIC      BIC   logLik deviance df.resid
##    184.8    207.9    -85.4    170.8      192
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5390 -0.2367  0.1427  0.2909  1.8719
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  id     (Intercept) 5.765    2.401
## Number of obs: 199, groups:  id, 78
##
## Fixed effects:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -2.72189    1.11457  -2.442  0.01460 *
## postM            0.06738    0.10300   0.654  0.51296
## txtIntervention  3.42309    1.07794   3.176  0.00150 **
## agegroup25-34    2.25874    1.01277   2.230  0.02573 *
## agegroup35+      1.98025    1.38527   1.430  0.15286
## baselineGood     2.78128    0.98734   2.817  0.00485 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) postM  txtInt a25-34 agg35+
## postM       -0.517
## txtIntrvntn -0.614  0.068
## agegrp25-34 -0.594  0.022  0.396
## agegroup35+ -0.398 -0.002  0.206  0.390
## baselineGod -0.604  0.034  0.450  0.380  0.275
```

How are the interpretations different from the GEE model?